DOI:
10.1017/apa.2019.48

Document Version
Peer reviewed version

Citation for published version (APA):
Godman, M., Mallozzi, A., & Papineau, D. (2020). Essential Properties are Super-Explanatory: Taming Metaphysical Modality. *Journal of the American Philosophical Association* , *6*(3), 316-334. https://doi.org/10.1017/apa.2019.48

Journal of the American Philosophical Association

CAMBRIDGE UNIVERSITY PRESS

Essential Properties are Super-Explanatory: Taming Metaphysical Modality

Journal:	*Journal of the American Philosophical Association*
Manuscript ID	APA-19-0111.R2
Manuscript Type:	Article
Keywords:	Kinds, Essences, Metaphysical modality, Species

Marion Godman, Antonella Mallozzi, & David Papineau

Essential Properties are Super-Explanatory: Taming Metaphysical Modality

Abstract

This paper aims to build a bridge between two areas of philosophical research, the structure of kinds and metaphysical modality. Our central thesis is that kinds typically involve *super-explanatory* properties, and that these properties are therefore metaphysically essential to natural kinds. Philosophers of science who work on kinds tend to emphasize their complexity, and are generally resistant to any suggestion that they have "essences". The complexities are real enough, but they should not be allowed to obscure the way that kinds are typically unified by certain core properties. We shall show how this unifying role offers a natural account of why certain properties are metaphysically essential to kinds.

Keywords:

Kinds, essences, metaphysical modality, species

Essential Properties are Super-Explanatory: Taming Metaphysical Modality

1 Introduction

Saul Kripke's examples of *a posteriori* necessities in *Naming and Necessity* revived the traditional distinction between the essential and accidental properties of things. Kripke was content, however, to leave this distinction on an intuitive basis. As a result, many philosophers still view metaphysically essential properties with suspicion, once the trivial cases of self-identity are put to one side (e.g. Sidelle 1989, Mackie 2006, Divers 2007).

In this paper we want to offer a principled basis for Kripke's distinction. In our view, *super-explanatory properties* lie behind all the non-trivial Kripkean *a posteriori* necessities. These are properties that explain the many shared features of things, and we argue that this makes them metaphysically necessary to those things. Super-explanatory properties thus provide a bridge between the actual world and the modal realm. Their status derives from their playing a certain causal role in the actual world, but because of this status they are held constant across all possible worlds.

In this paper we shall concentrate on the metaphysically necessary properties of natural kinds. But we believe the idea of super-explanatoriness applies more generally. At the end of the paper we shall briefly indicate how it also promises to account for the essential properties of particular objects like people and lecterns.

Philosophers of science who work on natural kinds have tended to emphasize their complexity and lack of sharp boundaries. Because of this, they are generally resistant to any suggestion that kinds have “essences”, viewing such thinking as grounded in outmoded metaphysics. (See e.g. Hull 1965, Sober 1980, Wilson et al. 2007, Häggqvist and Wikforss 2018.) We agree that natural kinds are complex, but will show that even so they are typically unified by certain super-explanatory core properties, which are thus naturally viewed as *essences*. While other philosophers have noted that causal mechanisms are needed to explain the unity of natural kinds (Boyd 1991, 1999, Craver 2000, Khalidi 2013), they have failed to appreciate how this role is typically played by single super-explanatory properties.

2 Kinds

Some categories have a rich structure, in that their instances share a great many properties.

For example, all horses are alike, not only in being horses, but in eating grass, growing manes, having uncloven hooves, and sharing a great number of other behavioural, anatomical and physiological features.

Following J.S. Mill, we shall call this sort of category a "Kind".[1] The defining characteristic of Kinds is that any Kind K will enter into a plurality of informative synthetic generalizations of the form: *All K are G*.[2]

Note that by no means all properties pick out Kinds. For example, square things, or soft things, or small things, do not form Kinds. There are no informative generalizations of the form: *All square/soft/small things are G*.

Higher biological taxa, as well as species, also form Kinds. For example, all mammals have fur, sweat glands, milk glands, and other distinctive anatomical features. Note how the features common to some higher taxon will be a subclass of those common to its subordinate taxa. The features common to all mammals are among the features common to all horses.

Biological categories are not the only Kinds. Chemical substances are also Kinds. All samples of sulphuric acid have the same melting and boiling point, the same

---

[1] As Mill put it "The class horse is a Kind, because the things which agree in possessing the characters by which we recognise a horse, agree in a great number of other properties as we know, and, it cannot be doubted, in many more than we know" (1843: 703-4).
[2] Our requirement for informativeness is that K should not logically entail G. Some Kinds are thinner than others, in supporting fewer generalizations. But any Kind supporting a plurality of generalizations is worthy of note. The generalizations associated with many Kinds, especially biological taxa and social categories, admit of exceptions.

propensity to combine with other substances in fixed proportions, the same liquid density, electrical and thermal conductivity, and so on.

Astronomical objects also fall into Kinds. All main sequence stars (those powered by hydrogen fusion) are spherical, in hydrostatic equilibrium, radiate energy of certain wavelengths, and so on. Other astronomical Kinds include red giant stars, white dwarf stars and supernovae. Further types of Kinds include meteorological categories and geological formations.

3 Kinds are Useful but not Anthropocentric

The rich structure of Kinds makes them very useful to human beings.

For one thing, we can normally ascertain that something is an instance of some Kind K by observing just one, or some few, of its characteristic Gs. For example, you can reliably identify horses by the characteristic shape of their heads, or gold by its characteristic electrical conductivity. And then, if you are appropriately informed, you can infer all the other characteristic Gs from this.

For another thing, there are often short-cuts to becoming appropriately informed about the Gs that are characteristic of some Kind. This is because Kinds fall into

Types that specify *which* properties will be shared by all the instances of Kinds of that Type.

So, for example, vertebrate species are one Type of Kind, in that all the animals in any given vertebrate species will share morphology, anatomy, physiology, diet, reproductive behaviour, and various other properties (but not injuries, or learned behaviours, for example.) Of course, the determinate morphology, anatomy, and so on, of any one vertebrate species will be different from that of the others; still, the different vertebrate species will all coincide in each being characterised by determinates of the same set of determinable properties.

Similarly, chemical substances are another Type of Kind. All samples of any given chemical substance in a given state of matter will typically have the same colour, taste, odour, density, electrical and thermal conductivity, and so on (but not, for example, the same shape, or size.)

If you know the "template" of determinable properties for a certain Type of Kind, as Ruth Millikan has put it (2000: ch. 1.8), then you are in a position to make *one-shot inductions*. From a traditional perspective, it can seem puzzling that one experiment on one piece of gold can tell you about the electrical conductivity of all gold, or that one dissection of a hippopotamus can tell you about all hippopotamus bladders.

However, once you know the template for chemical substances, or for vertebrates, then these inferences are secure.

All in all, then, facts about Kinds are very useful to humans who want to make inductive inferences. But it should not be thought that Kinds somehow come into existence in order to serve the inductive interests of humans. Rather, the Kinds first have their structure, and then humans use this structure to help them find their way around the world.[3] Most Kinds would exist even if no humans or other intelligent beings had ever evolved to base inferences on them. (Indeed it is likely that many of our distinctive intellectual powers evolved specifically to enable us to discern the pre-existing structure of Kinds.)

Later we shall have occasion to observe that entities like works of literature and social groups are also Kinds. Such entities do depend on humans for their existence. But the basic point remains. They aren't constituted in order to serve the inferential purposes of humans. It's just that their structure lends itself to these inferential purposes once they do exist.

4 Super-Explanatory Properties

[3] Of course, our interests will determine *which* Kinds we attend to, and moreover will often direct our attention to categories that are not Kinds at all (cf. Dupre 1993).

When we have a Kind K whose instances share many different properties G, there will typically be some single property E of their instances that causally explains this multiple commonality. Our eventual aim in this paper is to show that such "super-explanatory" properties are also essential in the modal sense of being possessed by Kind instances across all possible worlds. But that is for later. Our first task is to explain the role of super-explanatory properties in the actual world.

For example, the *atomic constitution* of gold explains why all samples of solid gold have the same density, electrical and thermal conductivity, melting and boiling point, and so on. More generally, the *molecular constitution* of any given chemical substance will explain why its instances share many corresponding properties.[4]

Similarly, the internal physical constitutions characteristic of the various Kinds of astronomical objects explain why the instances of such Kinds (the main-sequence

[4] More precisely, it is the microstructure consequent on molecular composition that accounts for the shared properties of chemical substances; accordingly, we shall understand citation of chemical formulae like "H_2O" as shorthand for the relevant microstructures; this is in line with the system of nomenclature developed by the International Union of Pure and Applied Chemistry (see Hendry 2016, Weisberg et al. 2019, esp. sect. 4.5). Some further complications: it is molecular *structure*, not chemical formula, that explains the properties of chemical compounds (isomers can be very different from each other); isotopy makes a difference to some (but not all) of the properties of elements and compounds; some chemical mixtures are Kinds in their own right. A rich literature explores the complexity of chemical and other natural Kinds (inter alia Hendry 2006, Needham 2008, Chang 2012, Khalidi 2013, Häggqvist and Wikforss 2018). We take this literature to show that there are more Kinds than you might initially suppose, often nested within each other, each with their own super-explainers for their shared properties.

stars, the white dwarfs, the red giants, and so on) share their many other common properties.

Moreover, as we shall see, the members of any given biological taxa will also share a single property that explains their multiple commonalities – though in the biological case, we shall argue, this will not be a shared intrinsic property, as it is with chemical or astronomical Kinds.

We can view such "super-explanatory" properties as illustrations of the Principle of the Common Cause. In general, when we find that some A and B are correlated (in the sense that they are *co-instantiated* more often than we would expect given their separate probabilities of occurrence), then it will be the case that either A causes B, or B causes A, or both A and B are joint results of some common cause.

Kinds are in effect rich complexes of correlations. All the many Gs characteristic of a Kind are correlated with each other. As soon as we know that something has one such G, this typically makes all the other Gs much more likely. However, at least in the cases we have considered so far, the different correlated Gs don't cause each other. The melting point of gold doesn't cause its density. The bladders of hippopotami don't cause their ears. So correlations between properties like these must be due to some single common cause, some super-explanatory property of the instances that is responsible for their all sharing so many other properties.

Some doubt that the Principle of the Common Cause can be defended in any general form (Arntzenius 2010). We disagree, but need not pursue the issue here. The more important point is that Kinds typically have super-explanatory cores, not whether this fact can be derived from some exceptionless principle.

So for us Kinds are any categories whose instances display multiple commonalities. A super-explanatory property for such a Kind is then a single property that causes all the other shared properties. Such causal powers will perforce be conditional, depending on laws of nature and possibly normal background circumstances, in addition to the super-explanatory property itself. Still, as long as a single property, plus such conditions, accounts for all the other shared properties of a Kind, we will count it as super-explanatory.

5 The Structure of Kinds

Do *all* Kinds have super-explanatory properties? *Functional* Kinds are one interesting case. Consider the Kind *aerial insectivore*. All the swallows, martins, swifts, insectivorous bats, and other flying insect-eaters share a range of properties, including acute sensory systems, ability to swoop, and pointed beaks or mouths. Kinds like this are the result of convergent evolution. The aerial insectivores all belong to species that have been shaped by similar selective pressures.

Some are suspicious of Functional Kinds, on the grounds that they only share a few superficial properties, by comparison with the multitude of properties shared by the members of any biological taxon.[5] While this is certainly true, and indeed a point to which we shall return below, we are happy to accept Functional Kinds as genuine Kinds, albeit of limited informativeness. In line with this, we would say that the common selective pressures that give rise to Functional Kinds are themselves a super-explanatory common cause for the shared properties of those Kinds.[6]

We don't completely want to rule out Kinds that don't in any sense have a super-explanatory core property. Maybe there are Kinds where the shared Gs *do* cause each other, rather than stemming from some common cause. Perhaps some meteorological categories are like this. (For example, *tropical cyclones* might be seen as occurring when strong winds, heavy rain, spiralling thunderstorms and a low-pressure centre all reciprocally reinforce each other; cf. Millikan 2017: ch. 1. Relatedly, psychiatric disorders like depression are arguably sustained by feedback between different symptoms; cf. Borsboom 2017.)

[5] Thus biologists typically attach more weight to *homologies* between traits with a common ancestry than to *analogies* resulting from common evolutionary pressures (Brigandt and Griffiths 2007; Godman 2015).

[6] Perhaps artefacts (like *cars*, as opposed to makes and models like *Vauxhall* and *Zafira*) also form Kinds, with their common features stemming from the shared intentions of their makers.

As with Functional Kinds, one might feel that there is something rather thin about Kinds like tropical cyclones, or depression, that gain their common properties from cycles of reciprocal causation. We shall not push this point, however. None of our arguments will depend on the claim that *all* Kinds have instances that share some super-explanatory property. Rather, we are interested specifically in analysing those Kinds that *do* have super-explanatory properties. If there are Kinds which lack such a property, so be it. They are not our focus in this paper.

On Richard Boyd's influential account (1991, 1999), natural kinds are "homeostatic property clusters". We certainly agree that Kinds are property clusters, in the sense that their instances share many properties. But we disagree with his emphasis on homeostasis in the sense of self-regulating feedback mechanisms. Perhaps the clustering of properties in a few special kinds—tropical cyclones, depression—is due to some such self-equilibrating homeostatic mechanism. But there is in general nothing homeostatic about the way that microstructures give rise to the shared properties of chemical substances. Nor will the account of biological Kinds developed in the section after next appeal to homeostasis in any way.

Points like these have led a number of authors to argue that Boyd's analysis should be replaced by a "simple casual account" of natural kinds (Craver 2009, Khalidi 2013). On this view, the clustering of properties in any natural kind will be due to some causal structure, but this need not involve any homeostatic process. We agree with this turn away from homeostatic mechanisms, but we feel that an undifferentiated appeal to causal structure misses the widespread significance of super-explanatory properties. In our view, the great preponderance of natural kinds owe their clustering of properties, not just to some causal structure or other, but to one single underlying property that serves as the common cause of all the other clustered properties.

For completeness, it is worth observing that some categories technically fit our definition of Kinds, in supporting a plurality of generalizations, even though there is *no* unified explanation for the correlations involved, not in terms of any causal structure. For example, jadeite and nephrite are traditionally lumped together as *jade*, on account of their similar colour, hardness and density, even though they are two quite different minerals (cf. Bird 2010). We shall say that a Kind is *genuine* if its correlations have a unified causal explanation, and *ersatz* when it results from two or more genuine Kinds fortuitously coinciding in some of their properties. Since ersatz Kinds are by their nature rare, depending on a lucky coincidence of properties in different genuine Kinds, we shall put them to one side in what follows. It might seem odd to count ersatz Kinds as Kinds at all, but nothing will hang on this in what follows, since our focus henceforth will be exclusively on Kinds who instances do share a single super-explanatory property.

To summarize our taxonomy, then: any category supporting a plurality of generalizations is a *Kind*; only those for which this clustering has a unified causal explanation are *genuine*; for most but not necessarily all genuine Kinds, that explanation involves a *super-explanatory* common cause.

6 Historical and Eternal Kinds

Consider a rather different type of Kind from those discussed so far – all the different copies of *Alice in Wonderland*. This qualifies as a Kind. The copies share a multitude of properties. Each copy tells the same story, with the same characters and scenes. Indeed they all tell it in the same way, with the same first word, the same second word, and so on, until the end.

The car model *Vauxhall Zafira* forms another Kind. All cars of this model have the same ingenious system of extra folding seats, the same undercar spare wheel holder, the same type of carburettor, and so on.

The correlations between the properties of these Kinds are not explainable by some common internal physical property, like molecular structure or stellar composition. Different copies of *Alice in Wonderland* can be physically quite different (paper, magnetic tape, and so on). And even if all *Zafiras* are physically similar, there is no one aspect of this similarity that provides a common explanation for all the other shared properties of *Zafiras*.

Rather, the super-explanatory feature shared by the members of these Kinds is their common origin. All the many instances of *Alice in Wonderland* have been reproduced in one way or another from Lewis Carroll's original manuscript. All the many *Zafiras* have been manufactured in line with the designer's original blueprint. That's why these instances all share a multitude of properties.

Ruth Millikan, from whom we have derived much of our general framework for thinking about Kinds, has coined the term "Historical" for Kinds whose shared properties are due to their instances all being copied from some common source (1999). She contrasts these with "Eternal" Kinds whose shared properties are explained by the instances having some common *intrinsic* physical property. (The term "Eternal" is perhaps not ideal, given that instances of Eternal Kinds can be in flux and short-lived – as with supernovae, say, or decaying elements – but we will stick with it.)

Literary works and car models might seem somewhat frivolous examples of Kinds. But consciously designed items like these are not the only Historical Kinds. Various types of social category can also be seen in this light: the many beliefs and behaviours characteristic of *Christians*, say, or *molecular biologists*, or *Japanese women*, are arguably shared because they are copied from prior models (cf. Godman 2105, forthcoming; Ásta 2018). In addition, as we shall now argue, biological taxa are also Historical Kinds.

7 Biological Taxa

What sort of Kinds are biological taxa? A first thought might be that they are Functional Kinds, with their shared properties accounted for by common selective pressures. But this fails to do justice to the rich structure of biological taxa. The members of any biological taxon typically share a wealth of features that are not explainable by the selective pressures that shaped that taxon. This is because natural selection does not design ideal organisms, but builds on what went before, is limited to materials thrown up by mutation, and is buffeted around by genetic drift. This ensures that biological taxa standardly share any number of non-functional features. We need to look beyond selective pressures to capture the nature of biological taxa.

Michael Devitt (2008, 2010) has recently argued that we need to recognize intrinsic[7] biological "essences", in the form of shared genomic material and associated developmental mechanisms, precisely in order to account for the many characteristic phenotypic commonalities displayed by the members of any given biological genus. (We have put "essences" in scare quotes to repeat the point that at this stage we are still not concerned with any modal issues, but simply with the explanatory role of genomic mechanisms in the actual world.)

[7] For our purposes it will suffice to understand an "intrinsic" property of an entity as one that is metaphysically independent of the possession of other properties by other entities.

Devitt's position might suggest that we should assimilate biological taxa to Eternal Kinds, taking their shared genomes and developmental mechanisms to play the role of super-explanatory intrinsic properties. However, this is not the only possible way of fitting biological taxa into our analysis. An alternative option is to view biological reproduction and ontogeny as a copying mechanism which produces new members of biological taxa from old ones. This would then render biological taxa as Historical Kinds.

Does this have to be an either/or issue? The examples offered so far might have suggested that Eternal and Historical Kinds will always be mutually exclusive. We have considered Eternal Kinds, like chemical substances and stars, whose instances have had no causal connection to each other, but have simply arisen whenever the conditions were right. By contrast, the instances of Historical Kinds are necessarily causally connected to each other, related by descent through copying.

Still, the definition of an Eternal Kind as having a common intrinsic core does not itself preclude its instances from being causally connected to each other. So it remains open that some Kinds could be *both* Eternal and Historical, in the sense that their instances share some common intrinsic property *and* reproduce by copying.

Biological taxa could be a case in point. We could view them both ways. We could focus on the genomes and developmental mechanisms, and view reproduction as simply a part of the process which creates items with that intrinsic nature. Or we could focus instead on the reproduction, and view the genomes and development as simply a part of the process by which new instances are copied from old ones. The two perspectives are obviously compatible. This suggests that biological taxa are akin to both Eternal and Historical Kinds.

We nevertheless want to argue that biological taxa should be counted as Historical and not Eternal Kinds. In an earlier paper (forthcoming), Godman and Papineau pointed out that (a) even in sexually reproducing taxa, a significant number of taxon-typical biological characteristics are not inherited through the sexual bottleneck via intrinsic genetic properties of the zygote, but via non-genetic influences like behavioural imitation, and (b) in single-celled non-sexual taxa like bacteria, genomic material does not play any super-explanatory role, but is just one of the features copied in mitosis. On this basis, Godman and Papineau argued that a uniform account of biological taxa should treat them as Historical rather than Eternal Kinds.

It now strikes us, however, that there is a more fundamental objection to viewing biological taxa as Eternal Kinds. It's not just that this position fails to deal with non-genetic inheritance and non-sexual reproduction. Rather, even if we stick solely to genetically inherited traits in sexual species, it never offers the right kind of super-explanation at all.[8]

Remember that we are looking for some common-cause explanation of the multiple correlations between the many characteristic phenotypic features of biological genera. However, the genomic material common to the members of a taxon will typically not be suited to play this role. This is because it will normally be a *conjunction* of different genetic properties, each one of which explains a different phenotypic feature. So we don't have *one* intrinsic property acting as a *common* cause for many phenotypic properties, but simply a list of different intrinsic genetic properties explaining different phenotypic properties. And this thus leaves us once more with an unexplained multiple correlation: why are all those different intrinsic genetic properties found together?

By way of comparison, note that it would not be much of an explanation of why all gold, say, has a distinctive colour, density, electrical and heat conductivity, and so on, to be told that gold contains an intrinsic property I_1 that determines its colour, a different I_2 that fixes its density, another I_3 for its electrical conductivity, and another I_4 for its heat conductivity, and so forth. This would give us no account of why all these intrinsic I-properties occur together in any sample of gold.

---

[8] We would like to thank Nicholas Shea for this objection.

By the same coin, it is no explanation of the many features common to all horses, say, to specify that horses all have some genetic material that determines manes, and some other genetic material that determines uncloven hooves, and some other genetic material that determines their head shape, and so on. We wouldn't yet have explained why all these different genetic features are found together in horses.[9]

In the case of gold and other chemical substances, of course, it is molecular structure that provides the requisite common cause. The different features of gold all stem from one common source, its atomic structure, and not from a number of different intrinsic properties. But in the biological case the different genetic elements in some taxon's genome are not tied together in this way.[10]

To find a property that explains why all the different phenotypic features of a biological taxon are instantiated *together*, we need to turn away from ontogenetic

[9] It might seem that this argument assumes a strong version of "bean-bag" genetics according to which genomes are composed of separate "genes" each of which independently determines some phenotype, when in truth different parts of genomes combine in complex ways in producing any given phenotype. Our argument only requires, however, that genomes display some level of modularity, and this much is widely agreed to be a prerequisite of evolution by natural selection.

[10] The trouble with taxon genomes as super-explainers is not that they are *conjunctions* of properties. Plenty of legitimate super-explainers, like molecular structures say, can be analysed as conjunctions of properties. Rather, the objection is that the conjunction of properties in a genome do not constitute a unified common cause that is responsible for each of the taxon's characteristics Gs, and so fails to identify the source of their correlation. We would like to thank Julien Dutant for pressing us on this issue.

development and view the taxon as a Historical Kind. The reason the phenotypic features of a taxon are found together is that its members are all descended from common ancestors who had those features. Biological reproduction is a copying process – offspring share the heritable features of their parents.[11] So, a set of properties that is conjoined in an ancestral population will be found together in the descendants, too.

Note that nothing in our analysis requires biological taxa or other Historical Kinds to be immutable. Reproduction does not have to mean exact replication; and natural selection operating on so-introduced and pre-existing variation can lead to changes in the set of common characteristics displayed by a Kind. When this happens to a significant degree, then this will amount to emergence of a new Historical Kind, as most familiarly illustrated by the emergence of new species and other biological taxa.

One last point before turning to modal matters. Ernst Mayr introduced a well-known distinction between *proximal* and *ultimate* biological explanations (1961). A proximal explanation tells us how an individual organism ontogenetically develops the

[11] In sexual taxa, offspring are only partial copies of their parents, due to sexual dimorphism and the sexual mixing of traits. We shall take these complications as read when we refer to the members of a sexually reproducing taxon as being "copied" from a common source. With clones like Dolly the sheep, on the other hand, we do have full copies of parents. Sometimes people wonder whether such clones properly belong to the relevant biological species. That depends on exactly what hangs on "species membership", but we would certainly count clones as belonging to the same Historical Kinds as the parents they are copied from.

phenotypic features that are characteristic of its species. An ultimate explanation tells us why that species phylogenetically came to have those characteristics in the first place. Proximal explanations will refer to the developmental program that enables the zygote to develop into a mature organism. Ultimate explanations will appeal to the historical circumstances that shaped the species, including selective pressures.

Devitt explicitly relates his analysis to Mayr's distinction. He observes that his intrinsic essences answer Mayr's proximal question, in that they explain individual development in terms of genetic material and associated developmental mechanisms. He also responds to Mayr's ultimate question, by adding a further historical element to biological essences. So Devitt's overall view is that taxon essences are partly intrinsic and partly historic: the intrinsic component explains proximal issues of individual development, while the historic component explains ultimate issues of taxon origins (Devitt 2010).

It is worth making clear that our analysis of biological taxa is motivated by a question that is quite distinct from both of Mayr's questions. Where he is interested in ontogeny and phylogeny, we are interested specifically in explaining why so many different properties are tightly *correlated* in the members of any given biological taxon, not in how those properties develop as those organisms mature from zygotes, nor in why the taxon came to have those properties in the first place.

It is clear enough that we are not addressing Mayr's first proximal question. When we explain the clustering of a taxon's properties in terms of the common ancestry of its members, we simply take developmental mechanisms for granted. We assume that offspring will resemble their parents, and use this to infer that all taxon members will share the properties that were originally displayed by the founding population. From our perspective, the development of mature organisms from fertilized zygotes is not explained, but rather presupposed as part of the copying mechanism that ensures the resemblance between parents and offspring (a resemblance, remember, that is differently accomplished in non-genetic inheritance and in non-sexual taxa).

Nor are we answering Mayr's ultimate question. Even though our biological super-explanatory properties are historical, they do nothing at all to address Mayr's ultimate issue. By way of analogy, note that when we explain why all the copies of *Alice in Wonderland*, or all the *Vauxhall Zafiras*, share so many correlated properties, in terms of their all being copied from a common source, this doesn't even start to explain why *Alice* and the *Zafira* were designed as they were in the first place. While these are certainly questions worth asking, they are not our present focus, and we do nothing to answer them. We simply take the features of the originals as given, and explain the commonalities in the copies by that alone. Similarly, when we explain why horses share so many common features, say, we start from the fact that some few original horses had those features, and explain the correlations by that alone, without wondering where those original features came from.

So our question is different from both of the questions traditionally addressed by biological theorists. They are interested in ontogeny and phylogeny, where we are interested in explaining correlations. Moreover, our approach simply assumes answers to the traditional questions, taking as given both the machinery that allows offspring to resemble their parents, and the original traits of founding populations.

Still, even if our question is not one that has traditionally concerned biological theorists, this does not mean it is not philosophically important in its own right.

In this connection, note first that our correlational question is the one we need to focus on if we want to uncover some notion of "core property" that applies to Kinds in general, and not only to biological taxa. All Kinds involve correlations that call for explanation in terms of super-explanatory properties. But they do not all have ontogenies or phylogenies about which to ask Mayr's proximate and ultimate questions. Chemical samples, for example, do not develop ontogenetically from seeds, nor do their shared properties result from evolutionary processes.

Moreover, as we are about to argue, it is specifically correlation-explaining super-explanatory properties that hold the key to metaphysical modality. If we want a species of "essence" that is invariant across metaphysically possible worlds, we need to focus on super-explanatory properties, rather than developmental machinery or evolutionary origins.

8 The Necessity of Super-Explanatoriness

We now turn to modal issues. A central thesis of this paper is that super-explanatory properties are metaphysically necessary.[12]

This proposal fits well with a range of common intuitions about which properties are metaphysically necessary to Kinds.

Take chemical substances. As we have seen, molecular structure plays the super-explanatory role for chemical substances. In line with this, there is a widespread intuition that chemical substances have their molecular structure necessarily. The intuitive consensus is that there are no metaphysically possible worlds where water is not H_2O, and moreover that, in any possible world, anything that is H_2O is water.

[12] Mallozzi 2018 also defends this assumption, and in particular explores its epistemological implications, focusing on examples of eternal Kinds. Kment 2014 is another who ties modal essences to explanation: however, for him this is part of a broader account of modality in terms of *metaphysical laws*, including ontological laws governing mereology and grounding; moreover, while Kment does appeal to causal considerations, he does not invoke the specific super-explanatory causal role that is our focus.

It works similarly with biological taxa and other Historical Kinds. Here the super-explanatory property is shared origin. Intuitions about necessary features are more disputed for biological than they are for chemical substances, but even so it seems natural enough to take origins to be essential to taxa: you can't possibly be a tiger unless you are part of the lineage that starts with the original tigers, and necessarily anything that is part of that lineage (without too many modifications) is a tiger. Corresponding intuitions apply to Historical Kinds in general. For example, you can't possibly be a copy of *Alice in Wonderland* if your provenance doesn't trace back to Lewis Carroll's original manuscript, and necessarily any reasonably faithful copy with such a provenance is a copy of *Alice in Wonderland*.

As we said, Functional Kinds tend to be less rich than other Kinds. But once more it is not unnatural to regard the selective pressures that super-explain their shared properties as modally necessary and sufficient for Kind membership. You can't possibly be an aerial insectivore unless natural selection has shaped you to catch insects, and necessarily anything so shaped is an aerial insectivore.[13]

We do not want to rest our case on intuitions, however. While we take our position to fit reasonably well with the intuitive judgements of both philosophers of modality and non-specialists, we attach more importance to the fact that our proposal also

[13] And artefacts will have the intentions behind them essentially. Necessarily something is a watch if and only if it has been designed to be a personal portable chronometer.

suggests a principled explanation for these intuitions. In our view, the necessity of super-explanatory properties is not just some arbitrary conceptual quirk, but a natural consequence of the role that metaphysical modality plays in our cognitive economy.

One important aspect of metaphysical modality is the way it structures counterfactual thinking. Many everyday concerns call for us to consider what would have happened under some counterfactual supposition. When we counterfactually suppose that some item lacks some property, we naturally hold most of its other actual properties fixed. (Suppose Nixon had lost the election . . .) But when we counterfactually suppose that some Kind lacks a super-explanatory property, we are prevented from holding most of its other properties fixed, given that counterfactually supposing away a cause typically requires us to suppose away its effects, too. So when we suppose away the super-explanatory core of some Kind, all bets are off. We are hypothesing away all the many correlated properties that distinguish that Kind from others. The natural reaction is that we have thereby supposed away the Kind altogether.

We thus see how super-explanatory properties *guide* and *constrain* counterfactual thinking. Without such constraints, counterfactual suppositions could be developed in virtually any way, and so become theoretically and practically irrelevant. Williamson 2007 similarly appeals to our "sense of how nature works" in reliable counterfactual thinking. While he doesn't speak in terms of essential properties, he stresses the role of background knowledge of "constitutive facts" for correctly developing the supposition in a counterfactual conditional. For Williamson, we can generally trust our capacity for assessing counterfactual conditionals to deliver the correct answers because it is informed and constrained by such a knowledge of "constitutive" facts. Our account here agrees with Williamson's and importantly further clarifies *what* those constitutive facts are —namely facts about essential super-explanatory properties—as well as *why* they have such a special status in counterfactual thinking (cf. Mallozzi ms.; Vaidya and Wallner 2018).

Note that nothing in our account rules out possible worlds where a liquid with the superficial properties of water (odourless, colourless, tasteless, and so on) is not H_2O, and ones where a species with the superficial properties of tigers (striped, solitary, carnivorous, and so on) are not descendants of the original tigers. But that liquid would not be *water*, nor that species *tigers*, precisely because their correlated properties do not stem from the same super-explanatory source as in the actual world. We don't count them as the same Kinds, despite their superficial resemblance, because they have been put together in a different way. They don't have the crucial property that pulls together the categories that exist in the actual world.

Before proceeding, it will be helpful to add some precision to our arguments. We have on occasion drifted into talking of super-explanatory properties of Kinds, and of Kinds possessing those properties in all possible worlds. But it is not strictly the Kinds themselves that have the super-explanatory properties, but their *instances*; and correspondingly what's strictly necessary isn't that the Kind has the essential property, but that any instance of the Kind does. Thus:

(1) (K)(E)(x)((K is a Kind, and E super-explanatory with respect to K) → nec (Kx → Ex)

There is also the converse implication:

(2) (K)(E)(x)((K is a Kind, and E super-explanatory with respect to K) → nec (Ex → Kx)

Principle (1) says that any super-explanatory E is *required* for Kind membership across possible worlds. E is an "essential property", in that you can't belong to the K if you lack it.

Principle (2) say that any super-explanatory E is *sufficient* for Kind membership across possible worlds. E constitutes an "essence", in that you must belong to the K if you have it.[14]

Note that it doesn't follow from (1) and (2) that a property that's essential to, or an essence of, some Kind must be essential to, or an essence of, the individuals that belong to that Kind. *Having professional accreditation* is essential to and the essence of *being a medical doctor* – necessarily you have the accreditation if and only if you are a doctor. But that accreditation is not essential to (and still less is it the essence of) individual doctors *being who they are*. Any doctor could have failed to be a doctor; indeed no doctor was always a doctor, and many doctors cease to be doctors.

9 Metaphysical and Nomological Modality

It has been put to us[15] that super-explanatory properties do not really involve *metaphysical* modality, but only highlight certain aspects of *nomological* modality.

[14] Many theorists who are suspicious of Kind "essences" take it as definitional that they must be *intrinsic* properties (Hull 1965, Sober 1980, Wilson et al. 2007, Rieppel 2010, Häggqvist and Wikforss 2018). Similarly, essentialists like Mark Ellis also view historical essentialism with suspicion, suggest that historical essentialists "creatively revive the illusion of species essentialism" (Ellis 2011: 670). Given our view of Historical Kinds, we obviously think a meaningful sense of historical essentialism can survive these critiques (see also LaPorte 2017, Godman and Papineau forthcoming).

[15] In particular by Graham Priest.

After all, super-explanatoriness depends on nothing more than nomological relations between properties in the actual world, and so doesn't have to be seen as taking us beyond actuality.

We agree that our account of Kind essences gives priority to the actual world. Super-explanatoriness is fully determined by actual-world nomological relations. But we don't think that this disqualifies it as an account of metaphysically modally essential properties. After all, super-explanatoriness involves a quite specific kind of nomological structure. It's not just a matter of one property being nomologically related to another, as gas pressure, say, is related to temperature. Rather, it involves a given property being the *common cause* of a *many* other properties, and thereby explaining why those other properties are so tightly *correlated together*.

Our proposal is thus not to *eliminate* modal Kind essences in favour of nomological necessity, but rather to reduce them to a specific kind of nomological structure. This structure is distinguished from causal-nomological structure more generally by featuring single common causes which give rise to many other properties in what we can call "one-to-many" causal networks.

We are proposing that, whenever we find this one-to-many nomological structure characteristic of a super-explanatory property, then that property will be possessed in all metaphysically possible worlds. We take this to explicate the familiar notion of the "nature" of things as the source of metaphysical necessity. In our view, we gain a better grip on this notion by equating it with the underlying cause of the many properties and behaviours that consistently co-occur in the instances of a Kind—namely, the essence of the Kind as we understand it.

To drive the point home, note that it is a consequence of our view that modal Kind essences would stay fixed even in possible worlds where the relevant laws of nature were different. Even if variation in laws of nature meant that H_2O were no longer odourless, colourless, tasteless, and so on, it would still be *water*. We might be explaining modal Kind in terms of actual-world super-explanatoriness. But even so our account clearly implies that these modal essences transcend nomological connections, by staying fixed even when the latter vary.[16]

In this sense, then, super-explanatory properties provide a kind of bridge between the actual world and the modal realm.[17] Super-explanatoriness depends on nothing beyond a specific kind of casual role in the actual world. But the specific structure of this causal role shows how metaphysical necessity is tied to the nature of things and

[16] As it happens, we are not committed to the contingency of laws of nature. But it enough for our point that modal connections would come apart from nomological ones *if* the latter were contingent.
[17] Vetter 2015 explores a different way of grounding modality in actual-world facts: her *potentialism* explains possibility in terms of potentialities of actual individual objects. For more on the relationship between potentialism and essentialist accounts of modality, see her "Essence and Potentiality" ms.

so explains why super-explanatory properties remain invariant across all possible worlds.

10 Identity and Super-Explanatoriness

It has been put to us[18] that the modal consequences of (1) and (2) follow directly from the necessity of identity, and therefore that super-explanatoriness in not needed to explain them.

The thought is that such identities as *water* = *H_2O*, and *tigerhood* = *being descended from the original tigers*[19], and so on, together with the necessity of identity, are themselves enough to ensure the necessary coextensiveness specified in the consequents of (1) and (2), without any help from the super-explanatoriness referred to in the antecedents.

We agree that the modal consequences of (1) and (2) follow from Kind identities like *water* = *H_2O*, and *tigerhood* = *being descended from the original tigers*, together with the necessity of identity. But we think that such Kind identities are themselves a

[18] In particular by Paul Horwich and Stephen Neale.
[19] Perhaps it would be better to frame these as generalized identities rather than as simple equalities between properties (cf. Correia 2010, Rayo 2013, Dorr 2016, Correia and Skiles 2019). And ideally we should be more explicit about the types involved: for example, by equating *being water* with *having the microstructure consequent on being composed of H_2O molecules*. Still, we trust that our simplifications will not affect our arguments.

consequence of (1) and (2). We identify water with H_2O specifically because this molecular structure is super-explanatory of the many shared properties of water. Because of this super-explanatoriness, we regard water and H_2O as necessarily coextensive, and this then leads us to identify them.

After all, there are theorists who don't agree that

(3) *water* = H_2O.

(Let us stick to water. Everything that follows in the section would apply equally to tigerhood.) These theorists reject an externalist account of the concept *water*, and instead hold that it refers to anything that satisfies such nominal requirements as *odourless, colourless, tasteless, freezes at* 0^{O}*C, etc*. They allow that all the actual water is in fact H_2O. But they see no reason why this property should have a privileged role in relation to the nature of the Kind *water*, at the expense of the properties by which we recognize the Kind. They thus maintain that

(4) *water = odourless, colourless, tasteless, freezes at* 0^{O}*C, etc.*

and deny that (3) *water* = H_2O.

In line with this, they hold that it is by no means necessary that water is H_2O, or indeed that H_2O is water. They think that there are possible worlds in which water—the odourless, colourless, etc. liquid—is not H_2O, and worlds in which H_2O is not water—the odourless, colourless, etc. liquid—due to different types of observers or different laws of nature.

Now, we of course think that these anti-externalists are wrong about water (and we would say the same about similar anti-externalist views about other Kinds). But the interesting question is *why* they are wrong.

Note that nothing is extensionally amiss with their account. They will classify just the same actual liquids as water as anybody else.

In our view, the reason for identifying *water* with *H_2O*, rather than with *odourless, colourless, tasteless, freezes at $0^{O}C$, etc,* is precisely that H_2O is super-explanatory with respect to water, in a way that the nominal features of water are not, despite their extensional adequacy. The identity (3) *water* = *H_2O* is thus a *consequence* of the earlier principles (1) and (2), not some independent claim from which they follow. Because H_2O is super-explanatory, we take it to be both necessary and sufficient for water across modal space, as in (1) and (2). And this is why we then identify water with H_2O, as in (3), and not with any set of its more superficial properties.

A super-explanatory property of a Kind plays a special role. It is a single property that pulls together all the many other common features of the Kind. None of the Kind's other nominal features have this role. They are simply effects of the common super-explanatory core.

This is why we take super-explanatory properties to be modally required for Kind membership, as in (1) and (2). An instance with a super-explanatory property has the feature that actually causes all the other distinctive properties of the relevant Kind. And so possible (or indeed actual[20]) instances still count as members of that Kind, even if they lack some of the nominal features of the Kind. Conversely, if we posit an instance that has all the nominal features, but not the super-explanatory essence, it doesn't count as a member of the relevant Kind. As we observed earlier, a liquid that was odourless, colourless, tasteless, etc., but was not H_2O, would not be water, despite its superficial resemblance. The category would have been put together by hand, so to speak, rather than deriving its shared features from the core property that pulls them together in the actual world.

So, yes, once we are given such identities as *water* = H_2O, and *tigerhood* = *being descended from the original tigers*, then it will follow that the properties that flank the

[20] In Kinds with generalizations that admit of exceptions, such as biological taxa, there will be real cases like this, like albino tigers without stripes. Here the knowledge that the instance still belongs to the Kind can inform our actual-world inductions as well as our counterfactual conclusions.

identity sign are necessarily co-extensional. But the reason for embracing these identities in the first place is precisely that the properties on the right-hand sides are super-explanatory essences.

11 Further Issues

Many believe, following Fine 1994, that necessary truth across all metaphysically possible worlds is never brute, but always a consequence of essences. While this fits our account of necessities that depend on super-explanatory Kind essences, we are uneasy about trying to force all metaphysical necessities into the single mould of essence.

For a start, there is the necessity of logic itself. This is not the place to explore its basis, but there is no obvious reason to suppose that this will hinge on the essence of anything. Nor do we think that the necessity of identity itself is helpfully viewed as depending on essences, given the way it follows from nothing but the legitimacy of *de re* modal constructions.

Having said this, we do think that the idea of super-explanatoriness promises to cast light on a wider range of metaphysical necessities than just Kind essences. Ruth Millikan has observed that *individual* people, animals, and other persisting objects are akin to Kinds in a number of respects. If we think of the "instances" of any such individual as its temporal stages, then these instances will share many distinctive properties, small subsets of which will typically suffice to identify the individual in question. Given this, we can view any persisting individual on the model of a historical Kind, with the shared features of its instances being super-explained by its origin, and elements of that origin—like the parents of people, or the material constitutions of lecterns—therefore coming out as essential properties.

It is worth noting how this further possible application of super-explanatoriness, while rendering certain metaphysical necessities non-brute, is at variance with Fine's progamme of basing all metaphysical necessity specifically in essences. Since different persisting individuals sometimes have the same origin—think of identical twins—we can have different individuals with all the same essential properties, which means that these essential properties do not constitute modally sufficient *essences* for those individuals.

Still, these are issues for another time. In this paper we hope that we have at least done enough to show that super-explanatoriness is the right way to understand the metaphysically necessary properties of Kinds.[21]

[21] We would like to thank Alexander Bird, Ned Block, Paul Boghossian, Michael Devitt, Julien Dutant, Jani Hakkarainen, Paul Horwich, Markku Keinänen, Muhammad Ali Khalidi, Boris Kment, Jaakko Kuorikoski, Jessica Leech, Tim Lewens, Matteo Mameli, Caterina Marchionni, Ruth Millikan, Stephen Neale, Daniel Nolan, Graham Priest, Jesse Prinz, Samuli Reijula, Jonathan Schaffer, Martin Smith, Nick Shea, Barry Smith, Michael Strevens, Anand Vaidya, and two anonymous referees for *the Journal of the American Philosophical Association* for helpful feedback and conversation.

References

Arntzenius, Frank (2010) 'Reichenbach's common cause principle'. *Stanford Encyclopedia of Philosophy (Fall 2010 Edition*), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2010/entries/physics-Rpcc/>.

Ásta (2018) *Categories We Live By*. Oxford: Oxford University Press.

Bird, Alexander. (2010) 'Discovering the essences of natural kinds'. in Beebee, H. and Sabbarton-Leary, N. *The Semantics and Metaphysics of Natural Kinds* (London: Routledge), pp. 125-38.

Borsboom, Denny (2017) 'A network theory of mental disorders'. *World Psychiatry,* 16, 5-13.

Boyd, Richard (1991) 'Realism, anti-foundationism and the enthusiasm for natural kinds'. *Philosophical Studies,* 61, 127-48.

Boyd, Richard (1999) 'Homeostasis, species, and higher taxa'. In Wilson, R. ed *Species: New Interdisciplinary Essays* (Cambridge: MIT Press), pp. 141-85.

Brigandt, Ingo and Griffiths, Paul Edmund (2007) 'The importance of homology for biology and philosophy'. *Biology and Philosophy,* 22, 633-41.

Chang, Hasok (2012) *Is Water H_2O?*. Dordrecht: Springer.

Correia, Fabrice (2010) 'Grounding and truth-functions'. *Logique et Analyse,* 53, 271-96.

Correia, Fabrice and Skiles, Alexander (2019) 'Grounding, essence, and identity'. *Philosophy and Phenomenological Research,* 98, 642-70.

Craver, Carl (2009) 'Mechanisms and natural kinds'. *Philosophical Psychology,* 22, 575-94.

Divers, John (2007) 'Quinean scepticism About de re modality after David Lewis'. *European Journal of Philosophy,* 15, 40–62.

Devitt, Michael (2008) 'Resurrecting biological essentialism'. *Philosophy of Science,* 75, 344-82.

Devitt, Michael (2010) 'Species have (partly) intrinsic essences'. *Philosophy of Science,* 77, 648-61.

Dorr, Cian (2016) 'To be F is to be G'. *Philosophical Perspectives,* 30, 39-134.

Dupré, John (1993) *The Disorder of Things: Metaphysical Foundations of the Disunity of Science.* Cambridge MA: Harvard University Press.

Ellis, Mark (2011) 'The problem with the species problem'. *History and Philosophy of the Life Sciences,* 33: 343-63.

Fine, Kit (1994) 'Essence and modality'. *Philosophical Studies,* 8, 1-16.

Godman, Marion (2018) 'Gender as a historical kind: a tale of two genders?. *Biology & Philosophy*, 33, 21.

Godman, Marion (2015) 'The special science dilemma and how culture solves it'. *Australasian Journal of Philosophy,* 93, 491-508.

Godman, Marion and Papineau, David (forthcoming) 'Species have historical not intrinsic Essences'. In A. Bianchi (ed.), *Language and Reality From a Naturalistic Perspective: Themes From Michael Devitt.* (Springer Editions).

Häggqvist, Sören and Wikforss, Åsa (2018) 'Natural kinds and natural kind terms: Myth and reality'. *British Journal for the Philosophy of Science,* 69, 911-33.

Hendry, Robin (2006) 'Elements, compounds and other chemical kinds'. *Philosophy of Science,* 73, 864-75.

Hendry, Robin (2016) 'Natural kinds in chemistry'. in Fisher, G. and Scerri, E. eds., *Essays in the Philosophy of Chemistry* (Oxford: Oxford University Press), pp. 253–75.

Hull, David (1965) 'The effect of essentialism on taxonomy–two thousand years of stasis'. *The British Journal for the Philosophy of Science,* 15, 314-26.

Khalidi, Muhammad Ali (2013) *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences.* Cambridge: Cambridge University Press.

Kment, Boris (2014) *Modality and Explanatory Reasoning.* Oxford: Oxford University Press.

Kripke, Saul (1980) *Naming and Necessity*. Oxford: Blackwell.

LaPorte, Joseph (2017) 'Modern essentialism for species and its animadversions'. In Joyce R. (ed.) *Routledge Handbook on Evolution and Philosophy* (Abingdon: Routledge), pp. 182-93.

Mackie, Penelope (2006) *How Things Might Have Been*. Oxford: Oxford University Press.

Mallozzi, Antonella (ms.) 'Essentialist constraints on counterfactual knowledge'.

Mallozzi, Antonella (2018) 'Putting modal metaphysics first'. *Synthese*, 1-20.

Mayr, Ernst (1961) 'Cause and effect in biology'. *Science,* 134, 1501-6.

Mill, John Stuart (1843) *A System of Logic, Ratiocinative and Inductive*. London: Parker.

Millikan, Ruth (1999) 'Historical kinds and the special sciences'. *Philosophical Studies*, 95, 45–65.

Millikan, Ruth (2000) *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.

Millikan, Ruth (2017) *Beyond Concepts*. Oxford: Oxford University Press.

Needham, Paul (2008) 'Is water a mixure? Bridging the distinction between physical and chemical properties'. *Studies in History and Philosophy of Science, Part A,* 39, 66-77.

Rayo, Agustin (2013) *The Construction of Logical Space*. Oxford: Oxford University Press.

Rieppel, Olivier (2010) 'New essentialism in biology'. *Philosophy of Science*, *77*, 662-73.

Sidelle, Alan (1989) *Necessity, Essence and Individuation*. New York: Cornell University Press.

Sober, Elliott (1980) 'Evolution, population thinking, and essentialism'. *Philosophy of Science,* 47, 350-83.

Vaidya Anand and Wallner, Michael (2018) 'The epistemology of modality and the problem of modal epistemic friction'. *Synthese,* 1-27.

Vetter, Barbara (2015) *Potentiality*. Oxford: Oxford University Press.

Vetter, Barbara (ms.) ‘Essence and potentiality’.

Weisberg, Michael, Needham, Paul and Hendry, Robin (2019) ‘Philosophy of Chemistry’, *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/chemistry/>.

Williamson, Timothy (2007) *The Philosophy of Philosophy*. Oxford: Oxford University Press.

Wilson, Robert, Barker, Matthew, and Brigandt, Ingo (2007) ‘When traditional essentialism fails: Biological natural kinds’. *Philosophical Topics,* 35, 189-215.